PoisonGAN: Generative Poisoning Attacks Against Federated Learning in Edge Computing Systems

Year: 2021 | Citations: 214 | Authors: Jiale Zhang, Bing Chen, Xiang Cheng, H. Binh, Shui Yu

---

Abstract

Edge computing is a key-enabling technology that meets continuously increasing requirements for the intelligent Internet-of-Things (IoT) applications. To cope with the increasing privacy leakages of machine learning while benefiting from unbalanced data distributions, federated learning has been wildly adopted as a novel intelligent edge computing framework with a localized training mechanism. However, recent studies found that the federated learning framework exhibits inherent vulnerabilities on active attacks, and poisoning attack is one of the most powerful and secluded attacks where the functionalities of the global model could be damaged through attacker's well-crafted local updates. In this article, we give a comprehensive exploration of the poisoning attack mechanisms in the context of federated learning. We first present a poison data generation method, named Data_Gen, based on the generative adversarial networks (GANs). This method mainly relies upon the iteratively updated global model parameters to regenerate samples of interested victims. Second, we further propose a novel generative poisoning attack model, named PoisonGAN, against the federated learning framework. This model utilizes the designed Data_Gen method to efficiently reduce the attack assumptions and make attacks feasible in practice. We finally evaluate our data generation and attack models by implementing two types of typical poisoning attack strategies, label flipping and backdoor, on a federated learning prototype. The experimental results demonstrate that these two attack models are effective in federated learning.